

# Photometric Variability in *Kepler* Target Stars. II. An Overview of Amplitude, Periodicity, and Rotation in First Quarter Data.

Gibor Basri<sup>1</sup>, Lucianne M. Walkowicz<sup>1</sup>, Natalie Batalha<sup>2</sup>, Ronald L. Gilliland<sup>3</sup>, Jon Jenkins<sup>2</sup>, William J. Borucki<sup>2</sup>, David Koch<sup>2</sup>, Doug Caldwell<sup>2</sup>, Andrea K. Dupree<sup>4</sup>, David W. Latham<sup>4</sup>, Geoffrey W. Marcy<sup>1</sup>, Soeren Meibom<sup>4</sup>, Tim Brown<sup>5</sup>

## ABSTRACT

We provide an overview of stellar variability in the first quarter of data from the *Kepler* mission. The intent of this paper is to examine the entire sample of over 150,000 target stars for periodic behavior in their lightcurves, and relate this to stellar characteristics. These data constitute an unprecedented study of stellar variability given its great precision and complete time coverage (with a half hour cadence). Because the full Kepler pipeline is not currently suitable for a study of stellar variability of this sort, we describe our procedures for treating the “raw” pipeline data. About half of the total sample exhibits convincing periodic variability up to two weeks, with amplitudes ranging from differential intensity changes less than  $10^{-4}$  up to more than 10 percent. K and M dwarfs have a greater fraction of period behavior than G dwarfs. The giants in the sample have distinctive quasi-periodic behavior, but are not periodic in the way we define it. Not all periodicities are due to rotation, and the most significant period is not necessarily the rotation period. We discuss properties of the lightcurves, and in particular look at a sample of very clearly periodic G dwarfs. It is clear that a large number of them do vary because of rotation and starspots, but it will take further analysis to fully exploit this.

*Subject headings:* stars: activity — stars: starspots — stars: statistics — stars: rotation — stars: solar-type — stars: variables: general — stars: late-type

---

<sup>1</sup>Astronomy Department, University of California, Berkeley, CA 94720

<sup>2</sup>NASA Ames Research Center, Moffett Field, CA 94035

<sup>3</sup>Space Telescope Science Institute, Baltimore, MD 21218

<sup>4</sup>Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138

<sup>5</sup>Las Cumbres Observatory Global Telescope, Goleta, CA 93117

## 1. Introduction

The *Kepler* mission provides an unprecedented set of data on the photometric variability of stars (Borucki et al. 2009). It is unparalleled in its combination of photometric precision and time coverage. It observes a large sample of stars, which consist primarily of main sequence stars in the solar neighborhood. Although the primary mission is to discover exoplanets, clearly the stellar lightcurves are themselves a rich source of astrophysical information. Stars can vary their luminosity for a variety of reasons. Our orientation in this paper is to concentrate on stellar magnetic activity (primarily seen through starspot modulation). Questions of interest in this area include the stellar rotation periods, the differential rotation as a function of latitude, the distribution of spot areal coverages, the distributions in longitude and latitude on different stars, the presence and distribution of active longitudes, the timescale for evolution of spots of different sizes, spot contrasts, and the behavior of all these as a function of stellar mass, age, and rotation. Of particular interest to the mission is the translation of photometric periods into stellar ages for planet-bearing stars. This paper is not intended to provide answers to any of the above questions, but rather to present the characteristics of the *Kepler* dataset and its potential for making progress on these questions. Starspots have long been studied with either photometry or spectroscopy (Strassmeier 2009). Ground-based coverage, however, tends to have many gaps, and the spot areal coverage has to be many times solar for a signal to be robust. The promise of solving these deficiencies was well-illustrated first by the MOST (Guenther et al. 2010a) and COROT (Affer et al. 2009) missions, but with less precision and length of coverage than will be available over the *Kepler* mission. We have discussed the general level of variability in the first quarter of *Kepler* data in our first paper (Basri et al. 2010); hereafter Paper I), where there is also a lengthier introduction to stellar variability in the context we are interested in here.

## 2. Observations and Data Reduction

The *Kepler* Quarter 1 (Q1) observations took place over  $\sim 33.5$  days between May 13 and June 15 2009. This paper refers to data gathered in Q1, which was almost entirely released to the public on June 15, 2010. We use only the so-called Long Cadence data which is formed on-board the spacecraft by summing 270 6 second integrations into successive 29.42 minute blocks (see (Jenkins et al. 2010a) for additional details). Q1 consists of 1639 such intervals from 13 May 2009 to 15 June 2009, which provide continuous coverage. 156,097 stars were observed, almost all of which belong to the sample of stars that are identified as the core exoplanet target list. The construction of this target list has been described

by (Batalha et al. 2010). As described in the Release Notes, the data product which is relevant to this paper is the so-called raw data (identified by the keyword `AP_RAW_FLUX` in the FITS files). This is not really raw pixel data - it has been added up in each target aperture, corrected for background, flat fielding, cosmic ray removal, non-linearities, and there is some treatment of selected instrumental effects and data gaps. What remains in the raw photometry are other instrumental artifacts (Caldwell et al. 2010), secular instrumental drifts (Jenkins et al. 2010a), and intrinsic astrophysical variability.

Subsequent steps in the *Kepler* pipeline attempt to render the data more suitable for the primary mission: the detection of transits. Currently the methodology employed (Jenkins et al. 2010b) is not suitable for a study of stellar variability; one cannot be sure what components of variability have been removed, and the effects of reduction are different for different amounts of stellar variability. By comparing raw and corrected lightcurves, we determined that only raw lightcurves are suitable as a starting point for work that concentrates in detail on the variability due to stars (that may change with future versions of the pipeline). We elect to use a form of differential photometry, produced by dividing each lightcurve by its median value then subtracting unity.

We next investigated whether the secular trends seen in almost all raw lightcurves have an instrumental component that can be easily distinguished from true stellar variability. We made a linear fit to each curve, and examined the slopes as a function of position within a given detector or across the focal plane. It is not surprising to find these slopes; the stars drift on the pixels due to pointing errors, positioning of the field, differential velocity aberration, and focal plane distortion due to thermal and other effects. The response of each pixel can vary. The psf of the *Kepler* photometer has been intentionally designed to spread a star over several pixels, and apertures are intentionally larger than a target psf (both help with photometric precision). Stars at the edge of each target aperture can contribute more or less light over time depending on the direction of motion. Because of the great precision of *Kepler* photometry, all these effects may be important, depending on the target star and neighboring star field being measured. Unfortunately, we conclude that to first order the overall effects on the measured intensity of a star are very local, and one cannot use the behavior of nearby stars to reliably remove trends in a given target. Of course, the linear trend in our fitting procedure could have a truly stellar component in some cases, and we have to acknowledge that in such cases it will be removed.

After removal (by subtraction) of linear trends, it is clear that for many quiet stars, there remain curvatures in the lightcurve over the month of Q1 (there is no reason that instrumental secular effects should be purely linear). These also are not correlated spatially with each other, but it is clear that there is an instrumental component to them, because it

is very common for such lightcurves with a linear fit removed to be high in the middle and low at the ends. With a quadratic fit removed, there was also some (though not as much) consistency in the slope of the curves at the ends. In most cases, therefore, we subtracted a third order polynomial from the lightcurve. This resulted in flat lightcurves for many of the quiet stars; again with the danger that in some cases an astrophysical signal with a timescale of a couple of weeks or longer has been removed. This means that we are likely missing some slower low-amplitude rotators.

We also tested the effect of removing a third-order fit from pure sine curves of 1 to 4 cycles, constructed using the same cadence and timespan as the real data. For a single cycle (a period of 33.5 days), the third order fit crosses the sine wave 5 times and remains close to it. After subtraction one is left with a much smaller amplitude, and the periodogram has its maximum peak at about half the original period. The corrected lightcurve is much less affected for a 2-cycle curve (with an intrinsic period of 16.75 days). The amplitude after subtraction of a polynomial fit remains the same (with a slightly altered shape for the corrected lightcurve), and the inferred primary period is closer to 16 rather than 17 days. The effects for more cycles are even smaller. We conclude that our reduction procedure only allows valid discussion of primary periodicities of 16 days or less. Obviously, as Kepler continues to collect data this issue will be re-visited, along with development of methods for joining data across time intervals when the spacecraft has left and returned to the target field.

A caveat to the removal of third order fits must be made in the case when a star is clearly quite variable primarily on a timescale of 10 days or more. That is to say, the corrections on quiet stars were never large in amplitude (not larger than about 10 times the rms fluctuations on short timescales), and essentially all of them were flattened with a third order polynomial. This indicates that the spacecraft does not generally produce signals with higher amplitude and higher frequency than the polynomial. We therefore made an exception to the removal of a third order fit in the following circumstance. After removal of the linear fit, the differential lightcurve can be analyzed for zero crossings (the number of times it goes from positive to negative or vice versa). For a quiet star with a flat lightcurve there will be many of these (albeit with small excursions from zero). For a star with substantial low frequency variability the number of zero crossings gives an indicator of what its timescale of variability is. We used this to avoid removing a third order polynomial in lightcurves first smoothed over 10 hours which then exhibit 10 or fewer zero crossings. This is clearly somewhat arbitrary, and could be refined in the future. It is worth noting that in the sine wave test described above, even the 4-cycle test would not have had a third-order polynomial removed, since it has too few zero crossings. Thus we believe that our procedure retains almost all of the variability we are primarily interested in.

Finally, there are a few cases (usually less than 5 per 1000) where the raw lightcurve contains a sudden large discontinuity. These are removed in the pipeline after the raw stage, but because we cannot use that version, we had to remove them ourselves. This was done by filtering all lightcurves with a routine that looked for large abrupt changes (a version of the routine we also use to look for eclipses and flares). This yielded few enough cases that they could be dealt with manually. It is still true, however, the best way to treat them is not always obvious, and we employed three different strategies. In most cases, the discontinuity occurred somewhere in the middle of a not too variable lightcurve. In that case we removed a third order polynomial on each side of the discontinuity; these tended to match across the gap, but if they did not we forced a match. This procedure removes variability on a shorter timescale than one polynomial fit across the whole lightcurve.

In other cases with more variable stars, it was clearly damaging to use the first procedure (the polynomials removed variations that are clearly stellar), so we simply removed the discontinuity by forcing a match across the gap, always adjusting the latter part of the curve to agree with the earlier part. This often resulted in a new residual linear slope, which we removed again. Finally, in some cases the discontinuity occurred near one end or the other, and we selected which of the above procedures gave a more plausible solution. Fortunately there were rather few discontinuities, although there remain some that are below our chosen detection threshold; these do not matter in the current analysis. This manual intervention consumes most of the reduction time; the rest of it (including making the measurements described in the next section) can be done in a few hours on a personal computer.

Clearly all of these corrections have the potential to alter an astrophysical signal. For now we must accept that the photometry is not perfect (nor will it ever be), and look forward to the much longer measurement of each star that *Kepler* will obtain, which should clarify in many cases what the right thing to do is. The purpose of this paper, in any case, is to produce a large statistical overview of the stellar photometry, rather than to model in detail particular stars. We are confident that the conclusions we draw do not depend too much on the details of the preliminary methodology we have employed in this paper. We are capturing most of the amplitude and characteristics of stellar variability on timescales less than about 2 weeks, and can distinguish between periodic and non-periodic behavior. We can see whether the observed behavior is repetitious over a month, and can detect flares and other short timescale phenomena.

### 3. Analysis of Periodicity, Variability, and Rotation

Once the data were reduced as described in Section II, we proceeded to collect certain measurements on each lightcurve. Here we describe only those which proved useful to the aims of this paper. As in Paper I, one useful quantity is the variability range  $V_{rng}$  (sometimes just called “range”), which is intended to represent the basic level of photometric variability. We slightly changed the means of computing this quantity from Paper I, taking the unsmoothed differential lightcurve, sorting it by differential intensity, then computing the range between the 5<sup>th</sup> and 95<sup>th</sup> percentile of intensity (this tends to avoid including really anomalous excursions up or down). We express  $V_{rng}$  in mmag for convenience (actually the units are differential intensity times 1085.84, which is nearly correct but not logarithmic like the actual magnitude scale). We determined the zero crossings in both the reduced lightcurve itself and when it is smoothed by 10 hours (20 points), tabulating both the number of crossings and the mean and median separation of crossings. For all segments between crossings (either above or below zero) we saved the mean and median of absolute maximum excursions and integrated fluxes. The high frequency noise  $HF_{rms}$  in the lightcurve was calculated by subtracting a four-point boxcar smoothed version of itself from the lightcurve and computing the standard deviation of the result.

We also computed a Lomb-Scargle periodogram for each lightcurve, using 400 points spanning a range of periods between one and 100 days. These parameters were chosen because the procedure uses a logarithmic period spacing that is denser at short periods, so we extend the period range beyond the data length to force sufficient sampling in the periods of interest. We do not make use of results for periods beyond half the data length (16 days) or for periods below 3 days (to concentrate on solar-type rotating stars). This is the most computationally intensive part of our analysis, and we rebinned the lightcurve to 2 hour bins to save time. It then only requires a few hours on a small computer to obtain periodograms for all 156,000 stars. We saved the position, height, and width of all the peaks in them. We collected the height and integrated strength of the highest two peaks, the shortest and longest significant periods, and the number of peaks at least 10% as strong as the strongest one.

#### 3.1. Samples of Periodic and Non-periodic Variables

Our first division of the data was made on the basis of the strength of the strongest peak in the periodograms. After extensive manual examination, we found that strengths  $P_{str}$  of 60 or more showed by eye a clear believable periodicity, those between 35 and 60 were marginal, and those below 35 did not appear to be period in an obvious way. The

values of  $P_{str}$  depend on the ratio of the periodic signal to the noise and on the number of data samples and periods tested. The formal false alarm probability for our case is 1% at  $P_{str}$  of 13, and absurdly low for  $P_{str}$  of 35, so we are being subjective but conservative in our judgment of what is a significant periodicity. The value of  $P_{str}$  for a pure sine wave is around 800. In Figure 1 we show the Kepler Input Catalog (KIC) stellar parameters for the group of stars with  $P_{str}$  above 60 and below 35 (we will refer to these as the periodic and non-periodic samples). There are features in the  $\log(g)$ - $T_{eff}$  plane that are due to artifacts in the way the KIC was constructed and in target selection criteria for the exoplanet search (Batalha et al. 2010). We checked that the distribution of crowding factors (the fraction of light in target apertures that doesn’t come from the intended target star) is very similar for the two samples; the median crowding factor is 0.135.

There are some clear differences between the two samples. Giants are much more populous in the non-periodic sample. As described by (Bedding et al. 2010) they tend to have complex variability; periodograms do not reach the level of  $P_{str}$  we have set as our minimum threshold for significant periodicity (and exhibit multiple period peaks of comparable strength). By contrast, there are many solar-type stars in both the periodic and non-periodic samples. Cool dwarfs tend to show up preferentially as periodic. Hot stars also show up in both cases; part of the reason for this is that we have not tested for very short periods. Many of the hot pulsators are not flagged in our diagnostics but would be if we extended the search to shorter periods. There are more than 60,000 stars in our periodic sample, and nearly that many in the non-periodic sample. The marginal stars make up the roughly 35,000 remainder. The large number of stars with detected periods means that *Kepler* will be quite powerful in measuring stellar rotation. We are only considering slightly over a month’s data here and more than a year has already been collected. On one hand we should be able to detect spot rotations both longer in period and perhaps lower in amplitude as more data is available. On the other hand substantial issues remain to be resolved about how well stellar variability can be distinguished from instrumental effects, and how to patch across data segments where the stars are on different pixels.

Figure 2 shows  $V_{rng}$  in the two samples as a function of effective temperature. For comparison,  $V_{rng}$  for the active Sun is about 1 mmag. The periodic sample has a cloud of points of high variability at all temperatures, and the non-periodic sample shows a two concentrations with lower variability. One is at temperatures corresponding to red giants (between  $4500K < T_{eff} < 5200K$ ) and the other corresponds to less variable solar-type stars. Most of the higher amplitude objects in the non-periodic sample are hot pulsators with periods less than a day, which are not flagged by us as periodic because we have filtered for longer periods to focus on stellar rotation.

### 3.2. Lightcurves of Periodic and Non-periodic Variables

Another way to look at the two samples is to examine the range of variability  $V_{rng}$  against the high frequency noise  $HF_{rms}$  in the lightcurve due to faintness of the stars. Figure 3 shows the comparison between them. Because of the way they are defined,  $V_{rng}$  tends to be at least 4 times  $HF_{rms}$  in almost all cases. This is the cause of the diagonal line along which many of the stars lie in both samples. These low-lying stars are ones for which the amplitude of variability is not much greater than the noise, although in the periodic sample a low-amplitude periodicity has been detected. Conversely, the stars lying well above the line show high-amplitude variability, whether it is periodic or not. Many of the giants show up rather clearly in the non-periodic sample at  $\log(V_{rng})$  between 0.2 to 0.5, and  $HF_{rms}$  between -1.2 and -0.7. As a reminder, in the units of  $V_{rng}$  the depth of an earth-sized transit is about -1.1 in the log. The objects which show up below the main diagonal are almost all cases where there are a few deep transit/eclipse dips. They don't have enough points to generate high periodogram power, and are essentially ignored by the range calculation (which drops the 5% extremes at each end), but have an influence on  $HF_{rms}$  (pushing it to the right off the main line). The periodic sample shows a much more robust cloud of higher amplitudes, at all values of  $HF_{rms}$ . Particularly for points above the main diagonal ridge, the lightcurves show obvious periodic variability, and in the upper ranges they almost all appear to be starspots rotating in and out of view (with a few eclipsing systems with ellipsoidal lightcurves also present).

In Figure 4 we show the distribution of the ranges for the periodic (thicker line) and non-periodic samples. This conveys some of the same information as in Fig. 3, but shows more clearly that the periodic sample has many more stars with high-amplitude variability (as would be expected if starspots were the source of the variability). The distributions are much more similar at the low-amplitude end. Here the stars are quiet (don't show much variability); a little more than half of those are non-periodic. We have used spot modeling to convert an amplitude into a spot covering fraction, assuming spot contrast of about 0.6 (which may not hold over all spectral classes). This yields the result that  $\log(V_{rng})=0.3$  corresponds to a spot coverage of 1% of the stellar surface. A simple estimation using half-black spots gives about the same answer. The active solar coverage is about 0.5%, the peak of the distribution here is at about 1%, and more than half of the stars flagged as periodic have coverages well above 1%, extending up to 20% or more in the largest cases. Some of those extreme cases are not spots but ellipsoidal binaries; others are indeed very spotted stars (judging from the variations in the lightcurve from cycle to cycle, which binaries should not exhibit as readily).

We now present some typical lightcurves. In Fig. 5a are some obviously periodic G



dwarfs. The parameter space from which they are randomly chosen is  $5800\text{K} < T_{\text{eff}} < 5850$ ,  $\log(g) > 4.2$ ,  $8 < V_{\text{rng}} < 12$ ,  $100 < P_{\text{str}} < 120$ , and period (days) between 4 and 12. Fig. 5b shows a set of M dwarfs with the same parameters (except  $3900\text{K} < T_{\text{eff}} < 4000\text{K}$ ). One can see from the noise which stars are fainter (the top middle M dwarf is substantially brighter than the others). There is not an obvious difference between the G and M stars, but of course we’ve chosen them from the same relevant parameter spaces (although it is also true that the G and M dwarfs don’t populate these spaces very differently). In Fig. 5c we return to the G dwarfs, but now look in the region  $2 < V_{\text{rng}} < 4$  and  $65 < P_{\text{str}} < 70$ . The intensity scale has been decreased by a factor of 2 to better display the decreased range. In most cases the periodicity is still obvious. In a couple of instances it is harder to pick out a periodicity by eye. Finally, we present some lightcurves of stars in the giant domain in Fig. 5d. These have  $4700\text{K} < T_{\text{eff}} < 4750\text{K}$ ,  $2.5 < \log(g) < 3$ ,  $2 < V_{\text{rng}} < 4$ , and  $20 < P_{\text{str}} < 25$ . They look different from the G dwarfs in that the variability is aperiodic and fairly consistent in amplitude.

We investigated whether one can isolate giants purely by their photometric behavior. It has been well established that they have a characteristic photometric signature (Gilliland 2008); the question we are asking here is how well one can eliminate dwarfs which might also show similar behavior (even though the vast majority of them do not). We provide one reasonably successful attempt to do this without using KIC gravities as priors for G,K,M stars (we use them to check afterward). We choose stars with  $V_{\text{rng}}$  between 1 and 5 (motivated by where they are seen in Fig. 3), and with the number of peaks within 10% of the primary peak power greater than 10 (based on the periodograms of known giants). Finally we pick the number of smoothed zero crossings to be greater than 40 (to pick out the right timescale for giant photometric excursions) but the number of unsmoothed zero crossings to be less than 200 (to eliminate noise as the primary source of zero crossings). Figure 6 shows the result (with the unsmoothed zero crossings as the ordinate and KIC gravity as the abscissa). Most of the giants pass through our filter while very few dwarfs do. The remaining dwarfs have lightcurves that indeed closely resemble giant lightcurves, and perhaps some of them are actually misclassified giants. This can only be resolved through ground-based spectroscopy, or if the promise of detecting all the dwarf parallaxes from *Kepler* data itself is realized.

### 3.3. Periodic Spotted Main-sequence Stars

We now take a look at the periodic sample in more detail. Figure 7 shows the distribution of primary periods within a sample that has been further restricted by choosing only stars with KIC gravities greater than  $\log(g)=4$  (which still leaves over 50,000 stars in the sample).

The distribution of periods rises towards 2 weeks. Beyond that we do not trust the results because the time coverage is only 34 days, and because removing a third order fit can definitely have an influence on whether the lightcurve shows features timescales longer than 2 weeks. The small peak at about 3.5 days might possibly be due to the fact that one of the guide stars was a spotted star with this period (see the Kepler Data Release 5 Notes; it was removed after Q1) which caused the pointing to vary periodically. The structure at longer periods is due primarily to the fact that the sampling of periods tested is becoming sparser (the actual sampling is given by the row of cyan plusses in the upper histogram). We also investigated whether the range of variability is a function of the primary period. Generally, shorter periods indicate more active stars (Pizzolato et al. 2003), and these might be both more obviously periodic or show larger amplitudes of variation (unless the activity were too uniformly distributed). We see that the bulk of the stars have  $V_{rng}$  a little greater than the active Sun, and there is modest evidence of greater variability at shorter periods (the lowest  $V_{rng}$  are found mostly at the long periods). Of course, by restricting this study to stars rotating fast enough to show periodicity in two weeks, we have biased the result against the slower rotators (many of which presumably currently lie in our non-periodic sample). It is known that activity saturates at short periods and that might be playing a role here. Still, it is perhaps a little surprising that there is not a more obvious effect, especially since we are sampling rotation periods up to solar in the case of a half-period of two weeks.

We remind the reader that we did not analyze periods below 2 days; there is indeed a population of short period cases but we expect that those are not due to rotation (except perhaps in close binaries). It is important to note that one cannot assume that the dominant period is the rotation period of the star, even when the signal looks very much like starspots. In many cases it is more likely to be the half-period; when a star has an asymmetric spot distribution each face produces a dip of a different depth so that there could be two major dips per rotation. If there is essentially one active location, or a continuous enough distribution around the star, two discrete dips may not appear and then the periodogram will find the actual period.

To produce an active Sun sample for comparison, 20 segments of SOHO DIARAD (Dewitte et al. 2004) data from 2001 were recast to have similar cadence and length as the *Kepler* data (as in Paper I). Periods are found between 10-50 days with a mean and median of about 20 days (and mean  $P_{str}$  of 200). Of course, most of these periods (even those matching the actual solar period) would not be considered valid in our current analysis (since they are derived from 33 day data segments and we don't trust periods over 16 days). We composed a similar sample of quiet Sun data from 1996 SOHO measurements, and actually obtained a similar period distribution with a mean  $P_{str}$  of about 150. The high significance of  $P_{str}$  is primarily due to the very low noise in the SOHO data. We performed the additional

experiment of removing a third order fit from the quiet Sun data; this reduced  $P_{str}$  to a little below 100 but left the dominant period distribution as rather similar.

In order to better study this, we selected a subsample out of the periodic sample by restricting it to KIC gravities greater than  $\log(g)$  of 4.2, temperatures between 3800K and 6200K, periods between 3 and 16 days,  $P_{str} > 70$ , and  $V_{rng} > 2$ . The upper panel of Figure 8 shows some trend in this sample of these nearly 14,000 clear dwarf rotators for  $P_{str}$  to increase with  $V_{rng}$ , but there is a great deal of scatter along both axes. In this periodic dwarf sample there is a general decline in population to higher periodogram strengths, but all were chosen to show very clear periods. The analogous plot for the non-periodic sample (lower panel) is quite different, showing again that periodic stars tend to be more variable (although there is a region of low variability in common). Even these non-periodic stars lie primarily above the formal criterion for a significant period (which is at  $P_{str} \sim 15$ ). For solar-type main sequence stars, high variability very likely arises from starspots (which then exhibit periodicity through stellar rotation), and objects which don’t exhibit periods (or half-periods) as short as 2 weeks are not as variable. We often see possible signs of differential rotation in the variable stars (even with such a short span of data). Such signs are typically that there are two dips of different depth, and they can be seen to move out of phase with respect to each other. An excellent longer span example of this is shown by CoRoT-Exo-2b (Alonso et al. 2008). The dominant period can change as the spot distribution changes (although we don’t have long enough data segments here to really see that).

#### 4. Conclusions and Future Work

We find that the amplitude of variability is generally substantially larger for stars that are clearly periodic than for those that are not. Among the non-periodic variables, giants are the dominant constituent. We are able to select for giants based on their photometric properties alone with reasonable efficiency. The high amplitude non-periodic stars are almost all high frequency pulsators (too high for our tailored periodogram test to pick them up). For the periodic variables, the dominant period distribution rises to the long end of what we consider well-determined (up to two weeks). Since a substantial fraction of stars with dominant periods of two weeks may actually have rotation periods of a month, we are sampling into the pool of main sequence stars with solar rotation. Some of the stars with even longer periods must wait for more time coverage, and some of them are quiet enough to have been classified as non-periodic in the current study. It is too early to give a real rotational period distribution or make statements about the age distribution in this sample.

Perhaps the main conclusion of this preliminary overview of stellar variability as seen

by the *Kepler* mission is that this will be a very rich and unique dataset for studying the surfaces of stars. This study is preliminary both because the data reduction will require further refinement and mostly because we have only treated the first month of what is already more than a year (expected to become at least 3.5 years) of nearly continuous measurements at very high precision of roughly 150,000 stars. Well over half of the sample of main sequence solar-type stars show variability at a level beyond the noise, and many of them show apparently periodic behavior (60,000 more clearly and 34,000 somewhat marginally). Even restricting the sample to solar-type main sequence stars with strong amplitudes and periodicities produces a population of 14,000.

We have only looked by eye at a small minority of them so far. The stars whose variability is periodic but clearly not due to spots (pulsators and eclipsers are the primary other variables) are both far less numerous and can be filtered out (or selected) with good efficiency. While it is quite obvious in many cases that we are looking at starspots rotating in and out of view, evolving in strength, and differentially rotating at different latitudes, more work is needed to distinguish between these and other sources of variability in more ambiguous cases. We believe that the ambiguity of some of the lightcurves will be reduced or eliminated as longer time series become available. A lot of additional work is needed to model the spotted cases sufficiently well to extract all the information that is present. This will allow a qualitative leap in our understanding of magnetic activity on stars.

The authors wish to thank the entire Kepler mission team, including the engineers and managers who were so pivotal in the ultimate success of the mission. LW is grateful for the support of the *Kepler Fellowship for the Study of Planet-Bearing Stars*. GB thanks the NSF through grant AST-0606748 for partial support of this work. Funding for this Discovery mission is provided by NASA’s Science Mission Directorate.

*Facilities:* The *Kepler* Mission

## REFERENCES

- Affer, L., Micela, G., Favata, R., Flaccomio, E. 2009, AIP Conf. Ser. 1094, Cool Stars, Stellar Systems and the Sun (Melville, NY: AIP), 341
- Alonso, R. et al. 2008, *A&A*, 482, L21.
- Basri, G. et al. 2010, *ApJ*, 713, L155.
- Batalha, N.M. et al. 2010, *ApJ*, 713, L109.

- Borucki, W. J. et al. *Science*, 327, 977.
- Batalha, N. B. et al. 2010, *ApJ*, 713, L103.
- Bedding, T.R., et al. 2010, *ApJ*, 713, L176.
- Caldwell, D. A. et al. 2010, *ApJ*, 713, L120.
- Dewitte, S., Crommelynck, D., Jukoff, A. 2004, *J. Geophys. Res.* 109, A02102.
- Gilliland, R. L., 2008, *AJ*, 136, 566.
- Guenther, D. B. et al. 2008, *ApJ*, 687, 1448.
- Jenkins, J. M. et al. 2010, *ApJ*, 713, L87.
- Jenkins, J. M. et al. 2010, *ApJ*, 713, L120.
- Koch, D. et al. *ApJL*, in press.
- Pizzolato, N., Maggio, A., Micela, G., Sciortino, S., Ventura, P. 2003, *A&A*, 397, 147.
- Strassmeier, K. G. 2009, *A&A Rev.*, 17, 251

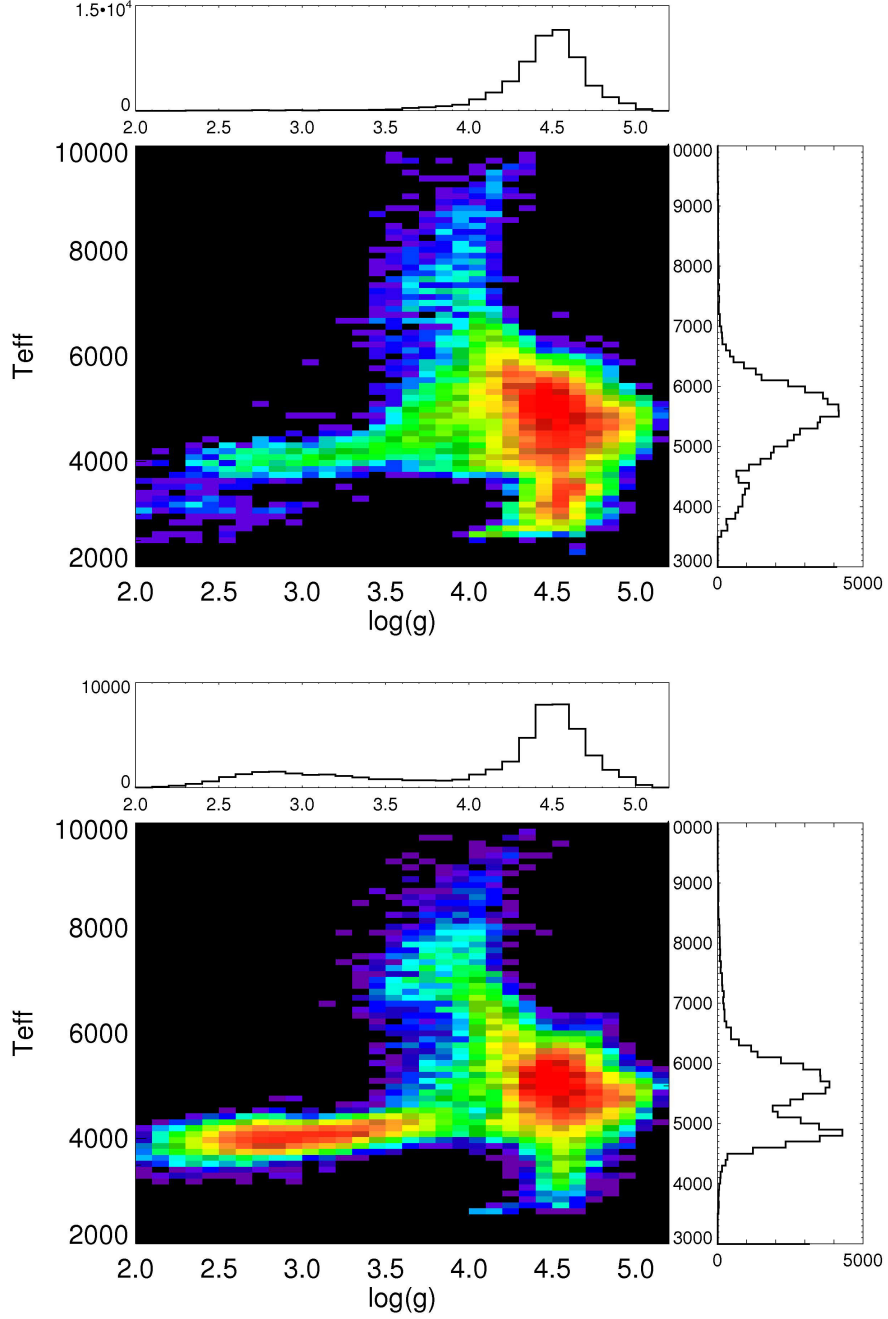


Fig. 1.—: Gravity vs. Temperature (using KIC parameters). The upper sample has stars with clear periodicity, while the lower sample is non-periodic. The periodic sample shows a stronger component of cool dwarfs. The non-periodic sample has post-main sequence stars (with low gravity); the hotter stars there tend to be pulsators with too high a frequency to trigger our filter for periods.

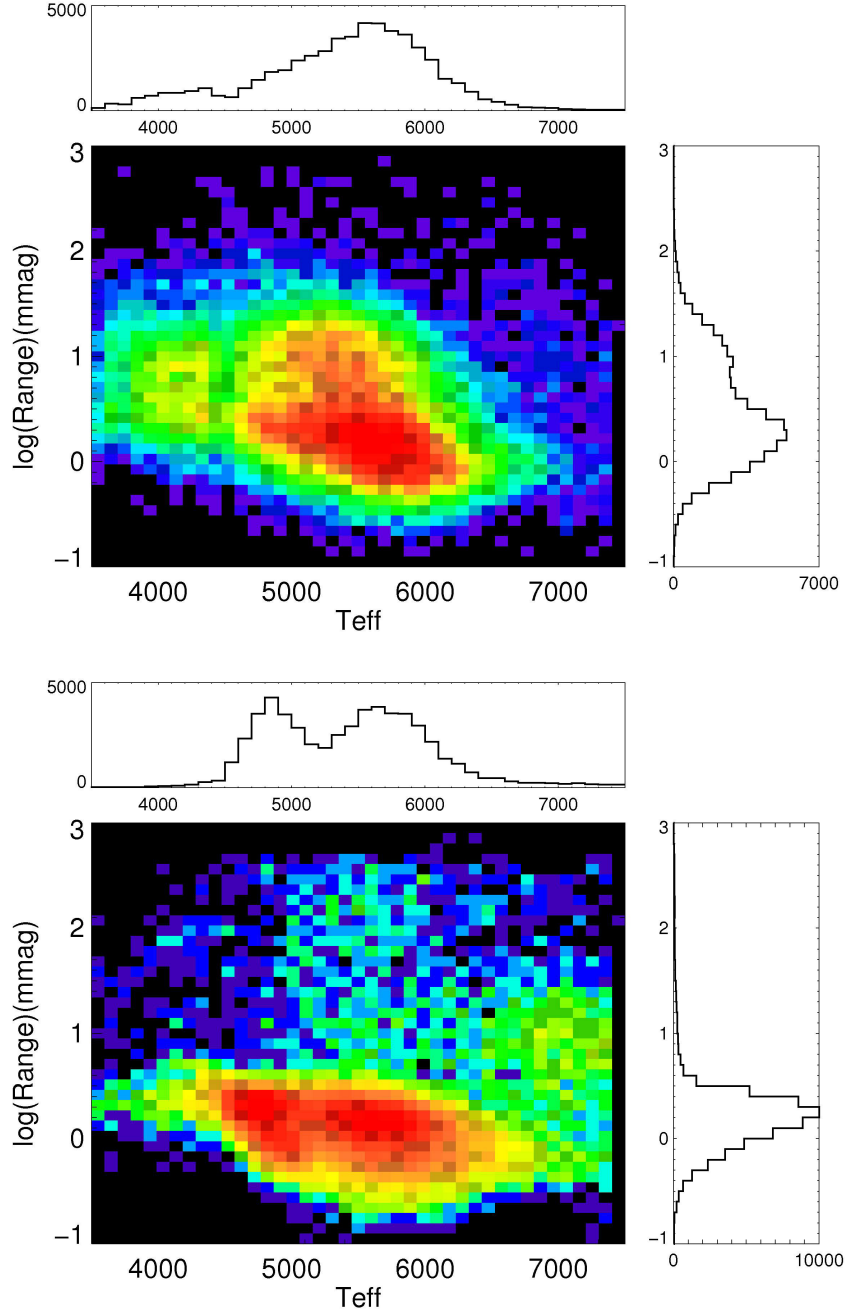


Fig. 2.—: Temperature vs. Amplitude of Variability (Range). As in Fig. 1, the upper panel shows periodic stars and the lower one non-periodic stars. The periodic stars have a cloud of higher amplitude objects at all temperatures, although there are many dwarfs with lower amplitudes for which as many are non-periodic. One can also see a tendency for cooler periodic stars to have greater amplitudes.

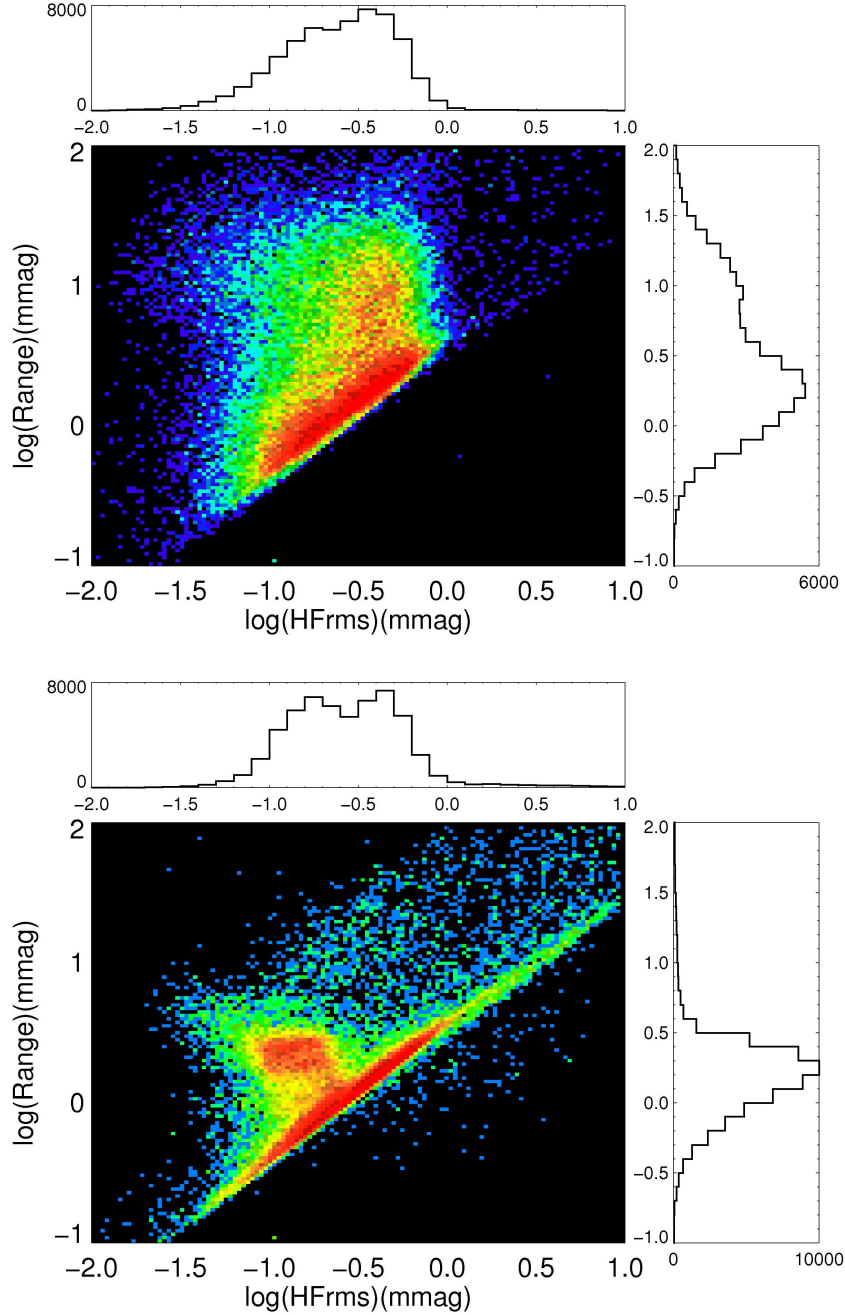


Fig. 3.—: High frequency noise vs. Variability Range. The abscissa is a measure of the photometric noise in the target (greater noise implies a fainter target). The lower diagonal ridge is populated by stars which do not have clear variability above the noise for both the upper (periodic) and lower (non-periodic) samples. Many more periodic stars show clear variability. The rectangular feature at  $\log(\text{Range}) \sim 0.4$  in the non-periodic sample is the locus of post-main sequence stars.



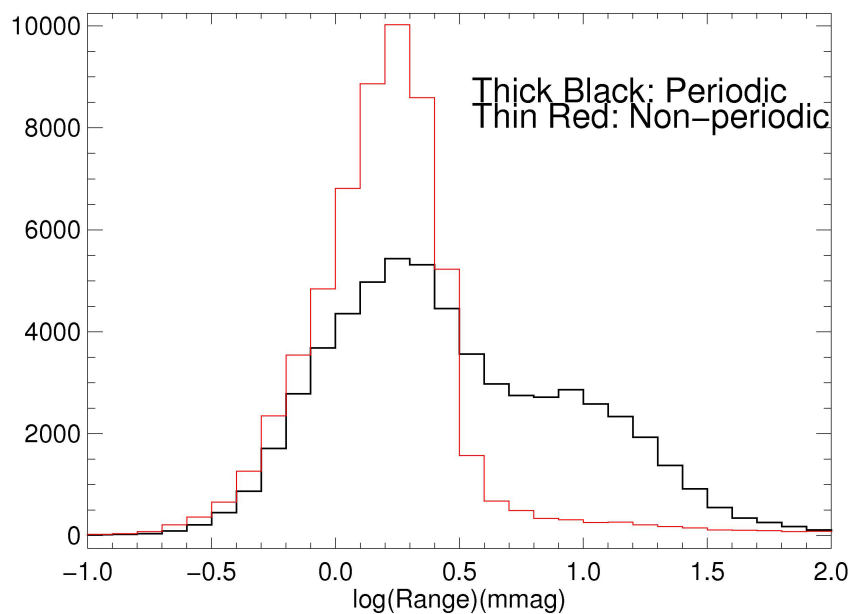


Fig. 4.—: Histograms of the amplitude of variability. The thick (black) curve is for the periodic sample of main sequence stars, and the thin (red) curve is for the non-periodic main sequence sample. This is a quantitative way to see the excess of variable stars in the periodic sample. There are similar numbers of stars with low variability in each sample (somewhat more are non-periodic).

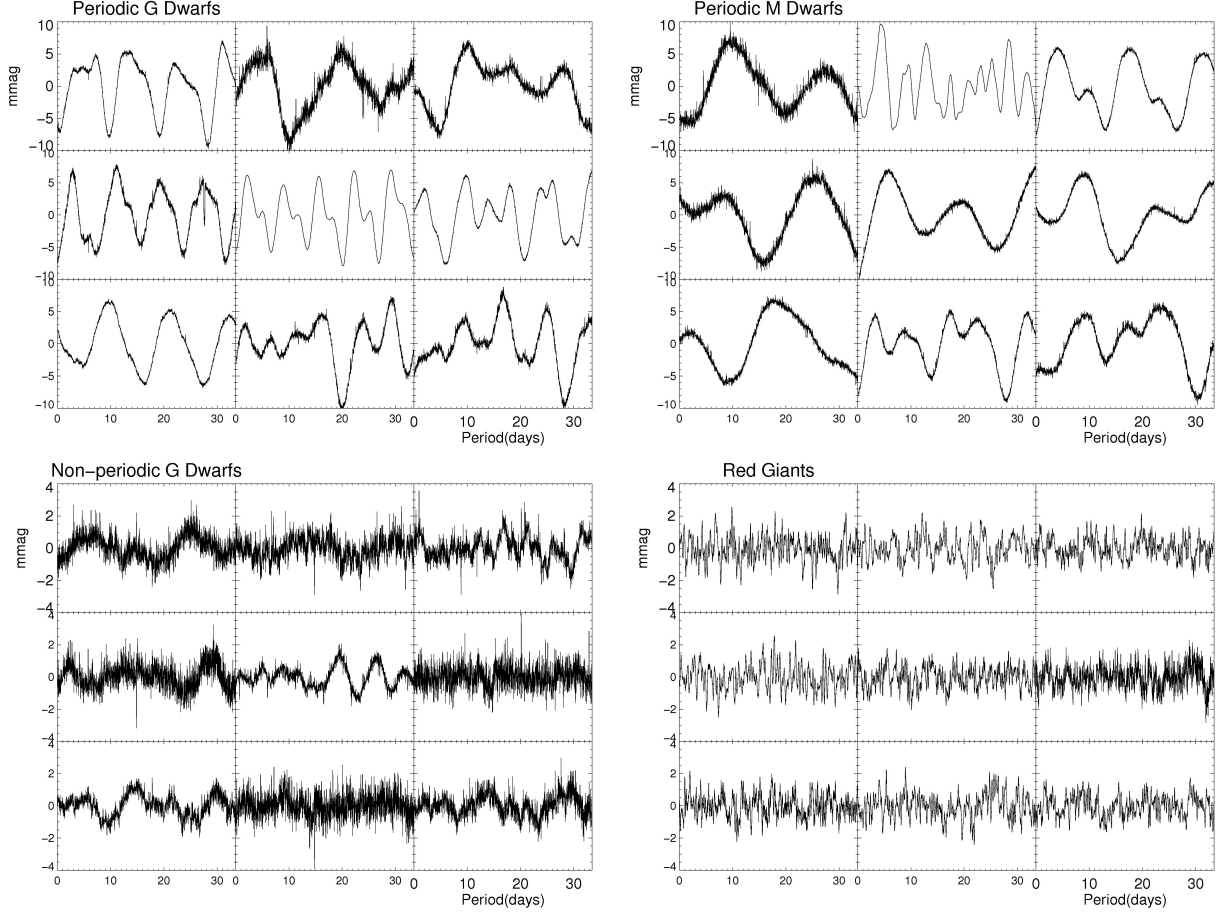


Fig. 5.—: Upper left: Lightcurves for clearly periodic G dwarfs (see text for a more exact description of stellar selection criteria). Kepler IDs for these stars are: 5302013, 9401951, 8367679, 11773022, 7200111, 4640983, 7037146, 12069336, 9269023. Upper right: Lightcurves for clearly periodic M dwarfs. Kepler IDs for these stars are: 11775907, 2424191, 4743351, 6382217, 6420895, 8558589, 9573685, 5950024, 4554367. Lower left: Lightcurves for weakly periodic G dwarfs. Note change of scale. Kepler IDs for these stars are: 9580212, 6125701, 8308260, 6837899, 10858832, 4366093, 4552939, 11598724, 3110216. Lower right: Lightcurves for red giants. Kepler IDs for these stars are: 12204548, 3427850, 10666932, 12208273, 4772722, 4725874, 10937855, 9783225, 4271855.

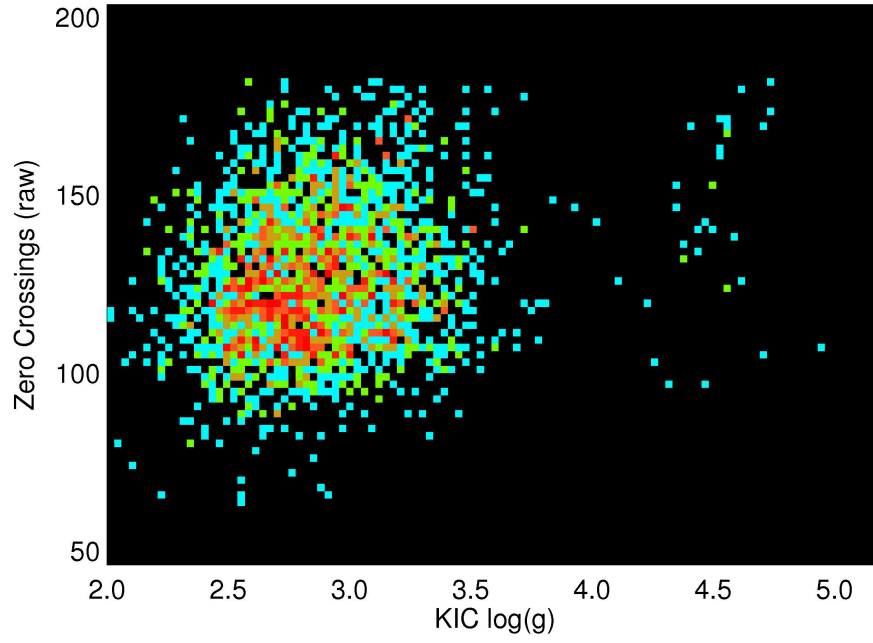


Fig. 6.—: The results of a photometric filter (see text) intended to select giants but not dwarfs. The ordinate is one filter that was employed (the number of times the differential lightcurve crossed zero). The KIC gravities were not used in choosing the stars, but afterwards to see whether giants were chosen. The high gravity cases look very much like giants photometrically (and it is possible that some were misclassified by the KIC).

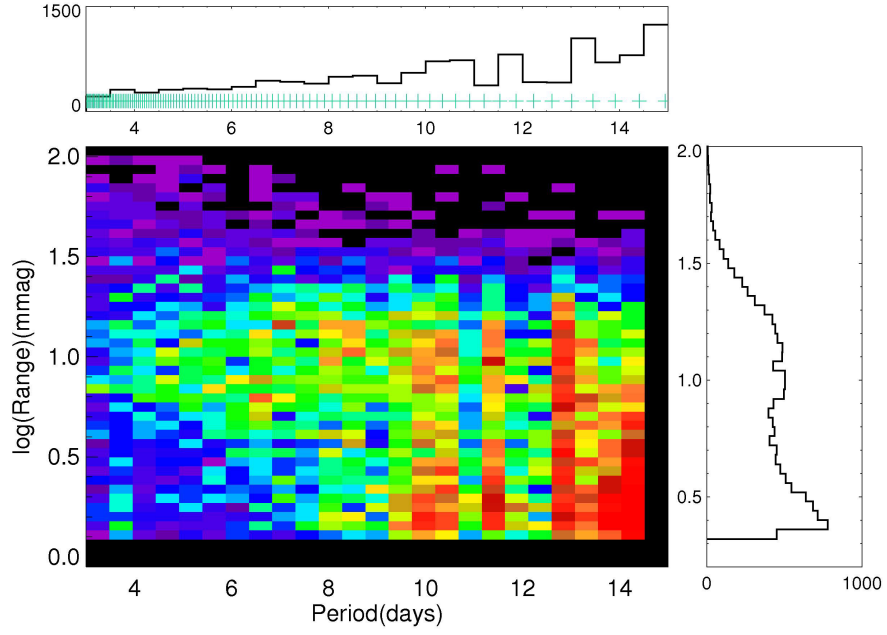


Fig. 7.—: The distribution of periods found for a selected sample of periodic main sequence stars (see text). The actual sampling of the periods search is given by the row of plusses in the upper histogram. The period histogram rises toward our limit of validity of 16 days. The distribution of variability range peaks just above the active Sun value, with a substantial tail at high variability (which extends to shorter periods).

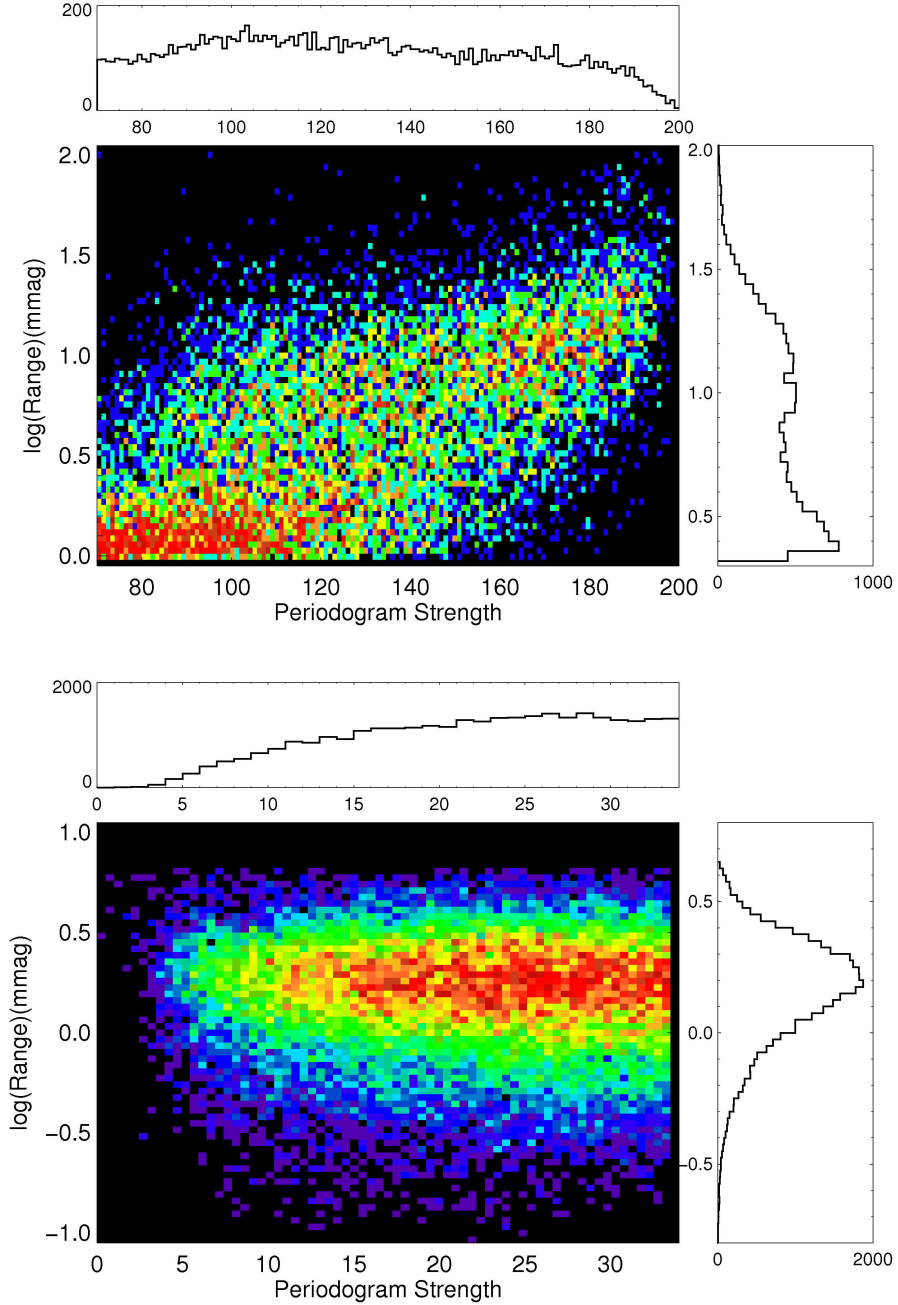


Fig. 8.—: Periodogram strength vs. variability amplitude in the periodic sample (upper panel). There is a clear tendency for these to be correlated, which is not surprising, but it does indicate that very large amplitude variations tend to be more strongly periodic. There is also quite a spread of amplitudes at a given value of periodogram strength. The lower panel shows the non-periodic sample. Note the change in scale for the abscissa. The amount of variability tends to be low, and there is no relation between it and what periodogram power is present (formally, values above 15 have false alarm probabilities  $< 10^{-5}$ ).